(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENT- UND MARKENAMT

(12) Übersetzung der europäischen Patentschrift

(97) EP 0 717 357 B1

(10) DE 695 29 728 T2

(51) Int. Cl.[7]: G 06 F 11/10
G 11 B 20/18

(21) Deutsches Aktenzeichen: 695 29 728.7
(96) Europäisches Aktenzeichen: 95 309 117.0
(96) Europäischer Anmeldetag: 14. 12. 1995
(97) Erstveröffentlichung durch das EPA: 19. 6. 1996
(97) Veröffentlichungstag der Patenterteilung beim EPA: 26. 2. 2003
(47) Veröffentlichungstag im Patentblatt: 20. 11. 2003

(30) Unionspriorität:
357847 16. 12. 1994 US

(73) Patentinhaber:
Hyundai Electronics America, Milpitas, Calif., US

(74) Vertreter:
Kahler, Käck & Fiener, 87719 Mindelheim

(84) Benannte Vertragstaaten:
DE, FR, GB

(72) Erfinder:
Stewart, John W., Wichita, US; Gates, Dennis E., Wichita, US; Dekoning, Rodney A., Wichita, US; Rink, Curtis W., Wichita, US

(54) Speicherplattenanordnungsgerät

DE 695 29 728 T2

Die vorliegende Erfindung betrifft Plattenarray-Vorrichtungen und insbesondere, aber nicht ausschließlich, einen Hardware-Plattenarray-Controller, der eine höhere Leistung und die Ausführung gleichzeitiger Datenbewegungen und Paritätsberechnungen bietet.

Eine redundante Anordnung kostenarmer Platten, die als RAID-Speichersystem bezeichnet wird, ist eine Sammlung von Plattenlaufwerken, die einem Host-Computersystem wie ein einziges großes Plattenlaufwerk erscheint. Ferner wird ein Teil der Plattenspeicherkapazität zum Speichern redundanter Informationen über Benutzerdaten verwendet, die auf dem Rest der Speicherkapazität gespeichert sind. Mit diesen redundanten Informationen kann die Plattenarray im Falle eines Ausfalls eines Elements des Array-Plattenlaufwerks weiter ohne Datenverlust arbeiten und erlaubt die Regenerierung von Daten auf ein Plattenlaufwerkselement einer Ersatzarray.

Mehrere RAID-Plattenarray-Designalternativen sind in einem Artikel mit dem Titel "A Case for Redundant Arrays of Inexpensive Disks (RAID)" von David A. Patterson, Garth Gibson und Randy H. Katz, University of California Report Nr. UCB/CSD 87/391, vom Dezember 1987 beschrieben. Der Artikel erörtert Plattenarrays und die Verbesserungen im Hinblick auf Leistung, Zuverlässigkeit, Leistungsaufnahme und Skalierbarkeit, die Plattenarrays im Vergleich zu einzelnen großen Magnetplatten bieten.

In dem Artikel werden fünf Plattenarray-Anordnungen beschrieben. Die erste RAID-Ebene umfasst N Platten zum Speichern von Daten und N zusätzliche "Spiegel"-Platten zum Speichern von Kopien der auf die Datenplatten geschriebenen Informationen. Schreibfunktionen auf RAID-Level 1 verlangen, dass Daten auf zwei Platten geschrieben werden, wobei die zweite "Spiegel"-Platte redundante Informationen erhält, d.h. dieselben Informationen, die auch auf die erste Platte geschrieben werden. Wenn Daten gelesen werden, dann werden sie von beiden Platten gelesen.

Systeme auf RAID-Level 3 umfassen eine oder mehrere Gruppen von N+1 Platten. Innerhalb jeder Gruppe werden N Platten zum Speichern von Daten verwendet, und die zusätzliche Platte dient zum Speichern redundanter Informationen, d.h. Paritätsinformationen. Bei Schreibfunktionen auf RAID-Level 3 wird jeder Datenblock in N Abschnitte zur Speicherung auf den N Datenplatten unterteilt. Die entsprechenden Paritätsinformationen werden dadurch berechnet, dass das Exklusiv-ODER-Produkt der auf die N Datenplatten geschriebenen Daten bestimmt wird, und werden auf eine dedizierte Paritätsplatte geschrieben. Beim Lesen von Daten muss auf alle N Datenplatten zugegriffen werden. Die Paritätsplatte dient zum Rekonstruieren von Informationen im Falle eines

Plattenausfalls.

Ein System auf RAID-Level 2 ist ähnlich wie die oben beschriebenen Systeme auf RAID-Level 3, beinhaltet jedoch zusätzliche redundante Platten zum Identifizieren von Plattenlaufwerksausfällen.

Systeme auf RAID-Level 4 setzen sich ebenfalls aus einer oder mehreren Gruppen von N+1 Platten zusammen, wobei N Platten zum Speichern von Daten verwendet werden und die zusätzliche Platte zum Speichern von Paritätsinformationen dient. Systeme auf RAID-Level 4 unterscheiden sich von Systemen auf RAID-Level 3 dadurch, dass zu speichernde Daten in größere Portionen unterteilt werden, die aus einem oder vielen Datenblöcken bestehen, die auf den Platten gespeichert werden. Schreibvorgänge erfordern weiterhin Zugriff auf zwei Platten, d.h. eine der N Datenplatten und die Paritätsplatte. Ähnlich erfordern Lesevorgänge typischerweise nur Zugriff auf eine einzige der N Datenplatten, es sei denn, die zu lesenden Daten überschreiten die auf jeder Platte gespeicherte Blocklänge. Wie bei Systemen auf RAID-Level 3, dient die Paritätsplatte zum Rekonstruieren von Informationen im Falle eines Plattenausfalls.

RAID-Level 5 ist RAID-Level 4 ähnlich, mit der Ausnahme, dass zusätzlich zu den Daten Paritätsinformationen über die N+1 Platten in jeder Gruppe verteilt werden. Jede Gruppe enthält zwar N+1 Platten, aber jede Platte beinhaltet einige Blöcke zum Speichern von Daten und einige Blöcke zum Speichern von Paritätsinformationen. Der Ort, an dem Paritätsinformationen gespeichert werden, wird durch einen vom Benutzer implementierten Algorithmus gesteuert. Wie in Systemen auf RAID-Level 4, so erfordern auch Schreibvorgänge auf RAID-Level 5 Zugriff auf wenigstens zwei Platten. Es verlangt jedoch nicht mehr jeder Schreibvorgang auf eine Gruppe Zugriff auf dieselbe dedizierte Paritätsplatte wie bei Systemen auf RAID-Level 4. Dieses Merkmal bietet die Möglichkeit, parallele Schreibvorgänge durchzuführen.

Wie bei RAID-Level 3, so können Paritätsdaten in einem System auf RAID-Level 4 oder 5 dadurch berechnet werden, dass ein bitweiser Exklusiv-ODER-Vorgang entsprechender Teile der über die N Datenlaufwerke gespeicherten Daten durchgeführt wird. Da jedoch jedes Paritätsbit einfach das Exklusiv-ODER-Produkt aller entsprechenden Datenbits von den Datenlaufwerken ist, lässt sich neue Parität leichter anhand der alten Daten und der alten Parität sowie der neuen Daten gemäß der folgenden Gleichung ermitteln:

neue Parität = (alte Daten XOR neue Daten) XOR alte Parität

Die in der obigen Gleichung gezeigte Paritätsberechnung für RAID-Level 4 oder 5 ist zwar weitaus einfacher als die Durchführung eines bitweisen Exklusiv-ODER-Vorgangs entsprechender Teile der über alle Datenlaufwerke gespeicherten Daten, aber ein typischer Schreibvorgang auf RAID-Level 4 oder 5 erfordert mindestens zwei Plattenlese- und zwei Plattenschreibvorgänge. Es werden mehr als zwei Plattenlese- und -schreibvorgänge für Datenschreibvorgänge benötigt, die mehr als einen Datenblock involvieren. Jeder einzelne Plattenlesevorgang involviert eine Suche auf und Rotation zu der/dem entsprechende(n) zu lesende(n) Spur und Sektor der Platte. Die Suchzeit für alle Platten ist daher das Maximum der Suchzeiten jeder Platte. Mit einem System auf RAID-Level 4 oder 5 ist daher ein erheblicher zusätzlicher Schreibaufwand im Vergleich zu einem Einzelplatten-Speichergerät oder mit Systemen auf RAID-Level 1, 2 oder 3 verbunden.

Um den Betrieb der Plattenlaufwerke mit einer Array zu koordinieren, um Lese- und Schreibfunktionen auszuführen, empfangene Daten auf den Plattenlaufwerkselementen der Array abzubilden, redundante Informationen zu generieren und zu prüfen und Datenwiederherstellung und -rekonstruktion bereitzustellen, sind komplexe Speichermanagementtechniken erforderlich. In vielen frühen Plattenarraysystemen wird die zum Durchführen dieser komplexen Speichermanagementtechniken notwendigen Array-Managementsoftware im Host-Computersystem ausgeführt. Das Hostsystem fungiert somit als Plattenarray-Controller und führt Erzeugung und Prüfung redundanter Informationen sowie die Koordination der zahlreichen anderen von der Plattenarray geforderten Speichermanagementvorgänge durch. Wenn der Host diese Funktionen ausüben muss, dann trägt dies erheblich zum Verarbeitungs-Overhead des Host bei.

Es ist Aufgabe der vorliegenden Erfindung, eine verbesserte Plattenarray-Vorrichtung sowie ein zugehöriges Verfahren zum Betreiben solcher Arrays bereitzustellen.

Gemäß der vorliegenden Erfindung wird ein Plattenarray-Controller für den Betrieb mit einem Hostcomputer und einer Mehrzahl von Plattenlaufwerken bereitgestellt, umfassend einen ersten schnellen lokalen Bus, eine Schnittstelle, die den genannten ersten schnellen lokalen Bus mit dem genannten Hostcomputer verbindet, wenigstens eine Schnittstelle, die die genannte Mehrzahl von Plattenlaufwerken mit dem genannten ersten schnellen lokalen Bus verbindet, und einen Prozessor, der mit dem genannten ersten schnellen lokalen Bus verbunden ist, um den Betrieb der mit dem genannten ersten schnellen lokalen Bus verbundenen Komponenten zu steuern, gekennzeichnet durch Paritätsassist-Logik, die direkt mit dem genannten ersten schnellen lokalen Bus verbunden

ist, ein lokales Datenspeichermittel, das mit dem genannten ersten schnellen lokalen Bus verbunden ist, und einen zweiten schnellen lokalen Bus, der die genannte Paritätsassist-Logik mit dem genannten lokalen Datenspeichermittel verbindet, wobei die genannte Paritätsassist-Logik die Aufgabe hat, Daten über den genannten zweiten schnellen lokalen Bus zu dem genannten lokalen Datenspeichermittel zu übertragen, um Paritätsdaten während des Plattenarray-Schreibvorgangs zu bestimmen.

Die vorliegende Erfindung stellt vorteilhafterweise neue und nützliche Plattenarray-Controller-Hardwarearchitektur bereit, die die Controllerleistung verbessert.

Die vorliegende Erfindung stellt auch eine Plattenarray-Controller-Architektur bereit, die gleichzeitige Datenbewegungen und Paritätsberechnungen sowie die parallele Ausführung von zeitunabhängigen Aufgaben zulässt.

Es ist ein weiterer Vorteil der vorliegenden Erfindung, dass ein Plattenarray-Controller bereitgestellt werden kann, der einzigartige Paritätsassist-Logik und einen dedizierten lokalen Speicher zum Ausführen von Paritätserzeugungs- und Datenrekonstruktionsvorgängen beinhaltet.

Ferner kann die vorliegende Erfindung vorteilhafterweise eine neue und nützliche Plattenarray-Controllerarchitektur bereitstellen, die es zulässt, dass gleichzeitig Datenblockbewegungen zwischen Speicher-E/A-Geräten und einem lokalen Speicher, Datenblockbewegungen zwischen einem Hostsystem und dem lokalen Speicher, Paritätsberechnungen und normale Array-Controller-Prozessor-Speicherabrufvorgänge ablaufen.

Es ist noch ein weiterer Vorteil der vorliegenden Erfindung, dass eine Plattenarray-Controllerarchitektur bereitgestellt werden kann, die auch einen Warteschlangenbetrieb von Blockbewegungen und einen Warteschlangenbetrieb für Paritätsaufgaben ermöglicht.

Es ist zu verstehen, dass sich die Erfindung auf das Plattenarray-Steuerverfahren bezieht, das in der obigen Vorrichtung ausgestaltet ist.

Es kann gemäß der vorliegenden Erfindung vorzugsweise ein Plattenarray-System zum Speichern von von einem Host-Computersystem empfangenen Daten und zum Übertragen von gespeicherten Daten zu einem Host-Computersystem bereitgestellt werden. Das Plattenarray-System beinhaltet einen ersten schnellen lokalen Bus; eine Schnittstellenschaltung, die den ersten schnellen lokalen Bus mit dem Host-Computersystem verbindet; eine Mehrzahl von Plattenlaufwerkselementen; wenigstens eine Schnittstellenschaltung, die die Mehrzahl von Plattenlaufwerkselementen mit dem ersten

schnellen lokalen Bus verbindet; eine Paritätsassist-Logikschaltung, die mit dem ersten schnellen lokalen Bus verbunden ist; einen lokalen Speicher; einen zweiten schnellen lokalen Bus, der die Paritätsassist-Logik mit dem genannten lokalen Speicher verbindet; und einen Prozessor, der mit dem ersten schnellen lokalen Bus verbunden ist, um den Betrieb der mit dem genannten ersten schnellen lokalen Bus verbundenen Komponenten zu steuern. Die Paritätsassist-Logik sendet Daten über den zweiten schnellen lokalen Bus zu dem lokalen Speicher, die für die Berechnung von Paritätsdaten bei Plattenarray-Schreibvorgängen erforderlich sind, und manipuliert die Daten, die auf dem lokalen Speicher gespeichert sind, um Parität bei Plattenarray-Schreibvorgängen zu ermitteln.

Die Erfindung kann auch eine Plattenarray-Steuervorrichtung für ein Plattenarray-System mit einer Mehrzahl von Plattenlaufwerkselementen zum Speichern von von einem Host-Computersystem empfangenen Daten und zum Übertragen gespeicherter Daten zu einem Host-Computersystem bereitstellen, wobei der genannte Plattenarray-Controller Folgendes umfasst: einen ersten schnellen lokalen Bus; eine Schnittstellenschaltung, die den genannten ersten schnellen lokalen Bus mit dem genannten Host-Computersystem verbindet; wenigstens eine Schnittstellenschaltung, die die genannte Mehrzahl von Plattenlaufwerkselementen mit dem genannten ersten schnellen lokalen Bus verbindet; eine Paritätsassist-Logikschaltung, die mit dem genannten ersten schnellen lokalen Bus verbunden ist; einen lokalen Speicher; einen zweiten schnellen lokalen Bus, der die genannte Paritätsassist-Logik mit dem genannten lokalen Speicher verbindet, wobei die genannte Paritätsassist-Logik Daten über den genannten zweiten schnellen lokalen Bus zu dem genannten lokalen Speicher sendet, die für die Berechnung von Paritätsdaten bei Plattenarray-Schreibvorgängen erforderlich sind, und die auf dem genannten lokalen Speicher gespeicherten Daten manipuliert, um Parität bei den genannten Plattenarray-Schreibvorgängen zu ermitteln; und einen Prozessor, der mit dem genannten ersten schnellen lokalen Bus verbunden ist, um den Betrieb der mit dem genannten ersten schnellen lokalen Bus verbundenen Komponenten zu steuern.

In der beschriebenen Ausgestaltung beinhaltet die Paritätsassist-Logik eine Paritätsassist-Maschine, einschließlich Exklusiv-ODER-Logik zum Errechnen von Parität, eine Dualport-Speicherschnittstelle, die die genannte Paritätsassist-Maschine mit dem genannten lokalen Datenspeichermittel über den genannten zweiten schnellen lokalen Bus verbindet, und eine Schnittstellenschaltung zum Verbinden der genannten Dualport-Speicherschnittstelle mit dem genannten ersten schnellen lokalen Bus. Der erste lokale Bus

ist ein schneller PCI-Bus. Der zweite Bus ist eine Dualport-Speicherschnittstelle.

Die Array-Controllerarchitektur ist skalierbar und unterstützt RAID-Betriebsarten 0, 3, 4 und 5. Die Architektur verfügt über eine Paritätsrechenmaschine mit hoher Bandbreite und eine gepufferte PCI-Schnittstelle, die mit der vollen Geschwindigkeit des schnellen lokalen PCI-Busses arbeitet. Sie verfügt über Multiprocessing-Fähigkeiten, so dass mehrere Aufgaben zur Ausführung in eine Warteschlange gereiht werden können. Ferner besteht die Möglichkeit, Blöcke von bis zu 128 MByte an zusätzlichem lokalem RAM-Speicher an den PCI-Bus anzuhängen. Der zusätzliche lokale Speicher hat zwei Ports, so dass PCI- und Paritätsvorgänge gleichzeitig ablaufen können.

Es ist somit zu verstehen, dass die Erfindung vorteilhafterweise eine skalierbare Hochleistungs-Hardwarearchitektur für ein Plattenarray-Speichersubsystem bereitstellen kann, das RAID-Betriebsarten 0, 3, 4 und 5 unterstützt. Die Architektur umfasst eine Paritätsrechenmaschine hoher Bandbreite, eine gepufferte PCI-Schnittstelle, die mit der vollen Geschwindigkeit eines PCI-Busses arbeitet, und einen dedizierten lokalen Speicher. Der dedizierte lokale Speicher hat zwei Ports, so dass PCI- und Paritätsvorgänge gleichzeitig ablaufen können. Die Architektur des Plattenarray-Controllers erlaubt Paritätsberechnungen und Speicherblockbewegungen ohne Störung des Controller-Prozessors oder seines assoziierten Speichers, so dass der Controller-Prozessor für die Verwaltung von Array-Tasksteuerung frei ist. Die Array-Controller-Konfiguration ermöglicht den gleichzeitigen Ablauf von Datenblockbewegungen zwischen Speicher-E/A-Geräten und lokalem Speicher; Datenblockbewegungen zwischen Host-SCSI-Verbindungen und lokalem Speicher; Paritätsberechnungen und normalen CPU-Speicherabrufen, Warteschlangenvorgängen für Blockbewegungen und Warteschlangenvorgängen für Paritätstasks.

Die Erfindung wird nachfolgend, jedoch nur beispielhaft, mit Bezug auf die Begleitzeichnungen näher beschrieben. Dabei zeigt:

Fig. 1 ein einfaches Blockdiagramm eines Plattenarray-Systems des Standes der Technik unter Verwendung eines schnellen PCI-Busses; und

Fig. 2 ein einfaches Blockdiagramm eines Plattenarray-Systems, das einen schnellen PCI-Bus und eine Hochleistungs-Paritätsfunktionsarchitektur gemäß einer Ausgestaltung der vorliegenden Erfindung verwendet.

Die meisten heute in Gebrauch befindlichen Plattenarray-Systeme sind in sich geschlossen und haben einen dedizierten Controller zum Ausführen der Array-

Managementsoftware, so dass das Host-System von diesen Arbeiten entlastet wird. Ein einfaches Blockdiagramm eines bekannten Plattenarray-Systems ist in Fig. 1 dargestellt. Das System beinhaltet einen intelligenten Array-Controller 100 zum Verwalten der Übertragung von Daten zwischen einem Host-Computersystem 12 und N Plattenlaufwerkseinheiten, von denen fünf als DRIVE A bis DRIVE E wie in Fig. 1 gezeigt identifiziert sind. In der Mitte des Array-Controllers befindet sich ein schneller lokaler Bus 102 wie z.B. ein Peripheral Component Interconnect (PCI) Bus. Eine Host-SCSI-Schnittstelle 104 und ein SCSI-Bus 14 stellen die Verbindung zwischen dem Host-Computersystem 12 und dem PCI-Bus 102 her. Ebenso sind die einzelnen Plattenlaufwerke DRIVE A bis DRIVE E über eine SCSI-Laufwerksschnittstelle mit dem PCI-Bus 102 verbunden, jeweils mit den Bezugsziffern 112A bis 112E identifiziert, und mit den entsprechenden SCSI-Bussen 114A bis 114E. Paritätsfunktionen werden durch eine Paritätslogikschaltung 108 und den lokalen Speicher 110 durchgeführt, die beide ebenfalls mit dem PCI-Bus 102 verbunden sind. Die Kommunikation zwischen den und der Betrieb der Controller-Komponenten wird durch einen Prozessor 106 gemäß in einem Prozessorspeicher 118 befindlichen Anweisungen gesteuert. Aufbau und Betrieb des in Fig. 1 gezeigten Array-Controllers, sowie die im Controller enthaltenen Komponenten, werden von der Fachwelt verstanden.

Der RAID-Speicherprozess erfordert viele Paritätsberechnungen und Datenbewegungsvorgänge zum Erzeugen der notwendigen Datenredundanz oder zum Rekonstruieren von Daten nach einem Plattenausfall. In der in Fig. 1 gezeigten und oben beschriebenen Array-Controller-Architektur wird eine starke Nutzung des PCI-Busses 102 benötigt, um neue Daten, alte Daten, rekonstruierte Daten, alte Paritätsinformationen und neue Paritätsinformationen zwischen dem Host-Computersystem 12 zu übertragen. Die Array treibt DRIVE A bis DRIVE E, den lokalen Speicher 110 und die Paritätslogik 108 zum Erzeugen neuer Paritätsinformationen bei einem Array-Schreibvorgang oder zum Rekonstruieren Daten nach einem Array-Ausfall.

Das in Fig. 2 gezeigte Plattenarray-System hat viele der in Fig. 1 gezeigten und oben erörterten Komponenten. Die Komponenten, die die Controller-Architekturen und Array-Systeme gemeinsam haben, erhielten in Fig. 1 und Fig. 2 dieselben Bezugsziffern. Gemeinsame Komponenten der in Fig. 1 und 2 gezeigten Systeme sind u.a.: schneller PCI-Bus 102, Host-SCSI-Schnittstelle 104, SCSI-Bus 14, SCSI-Laufwerksschnittstellen 112A bis 112E, SCSI-Busse 114A bis 114E, Plattenlaufwerkselemente DRIVE A bis DRIVE E,

Prozessor 106 und Prozessorspeicher 118.

Die in Fig. 2 gezeigte Array-Controller-Architektur verbessert jedoch die bisher bekannten Array-Controller-Architekturen, indem sie die Paritätslogik und den lokalen Speicher, die zur Paritätserzeugung und zur Rekonstruktion von Daten verwendet werden, vom PCI-Bus 102 entfernt. Die in Fig. 2 gezeigte Controller-Architektur bietet eine unabhängige Paritätsassist-Maschine 132 und einen lokalen Speicher 136, die durch eine Dualport-Speicherschnittstelle 134 miteinander verbunden sind. Eine schnelle PCI-Schnittstelle 130 ermöglicht die Übertragung von Daten zwischen dem PCI-Bus 102 und dem lokalen Speicher 136.

PCI-Busschnittstelle 130, Paritätsrechenmaschine 132 und Dualport-Speicherschnittstelle 134 sind auf einem einzigen RAID-Paritätsassist-Chip 128 zusammen integriert dargestellt. Diese drei Komponenten arbeiten jedoch unabhängig auf koordinierte Weise, um schnelle Datenspeicherung und Paritätsberechnungen durchzuführen und dabei ein Minimum an Host-CPU-Zyklen zu brauchen. Ein schneller lokaler Bus 138 verbindet den RAID-Paritätsassist-Chip mit dem lokalen Speicher 136.

Bei einem Plattenschreibvorgang werden Datenblöcke vom Host-SCSI-Kanal 14 empfangen und im lokalen Speicher 136 gespeichert. Die Paritätsassist-Maschine 132 liest die Daten, berechnet die Paritätsinformationen und schreibt sie zurück auf den lokalen Speicher 136. Die endgültigen Datenblöcke werden dann vom lokalen Speicher 136 auf die entsprechenden Plattenlaufwerke geschrieben.

Plattenlesevorgänge beginnen mit Daten, die von den Plattenlaufwerken auf den lokalen Speicher 136 bewegt wurden. Diese Daten werden dann auf den Host-SCSI-Kanal 14 übertragen, es sei denn, die Paritätsinformationen werden zum Rekonstruieren der Daten benötigt.

Da die Architektur äußerst speicherintensiv ist, ist der lokale Speicher 136 für einen Betrieb mit hoher Bandbreite mit einer 72 Bit breiten Schnittstelle organisiert. Es sind auch verschachtelte Zyklen für maximale Leistung möglich. Das Speichersystem kann mit der vollen Taktfrequenz des RAID-Paritätsassist-Chip arbeiten. Auch eine Auffrischsteuerung ist vorhanden, so dass der wirtschaftliche dynamische RAM zur Speicherung verwendet werden kann. Ein Abschaltmodus ist vorgesehen, um die Speicherdaten im Falle eines Stromausfalls zu schützen.

Das PCI-Schnittstellenmodul 130 kann mit vollen PCI-Bus-Übertragungsgeschwindigkeiten arbeiten. Es ist so ausgelegt, dass es auf alle

Busbetriebsarten von einem Byte bis zu Bursts bei voller Bandbreite reagiert. Das PCI-Modul enthält ein 128 Byte FIFO-Register zum Speichern von Daten, die vom PCI-Bus akzeptiert wurden. So kann die PCI-Übertragung auch dann fortgesetzt werden, auch wenn der lokale Speicher vorübergehend nicht zur Verfügung steht.

Die Paritätsassist-Maschine 132 bearbeitet Daten, die im lokalen Speicher 136 gespeichert sind. Sie berechnet Parität bündelweise unter Verwendung eines 128-Byte-FIFO zum Speichern der Zwischenergebnisse. Die Zwischenergebnisse werden nicht zurück auf den lokalen Speicher geschrieben. Zur Erzielung maximaler Leistung verwaltet die Steuerlogik für die Paritätsassist-Maschine Zeiger auf jeden benötigten Datenblock. Die Paritätsmaschine arbeitet mit der vollen Geschwindigkeit des lokalen Speicherbusses zur Bereitstellung der schnellstmöglichen Leistung für die Speicherbandbreite.

Die Paritätsassist-Maschine enthält vier separate Abschnitte, die es zulassen, dass zusätzliche Tasks in eine Warteschlange eingereiht werden, während eine Task ausgeführt wird. Jede Task führt ihr eigenes Steuer- und Statusregister, so dass durch Taskplanung keine gerade ausgeführte Task gestört wird. Es stehen mehrere Konfigurationsoptionen zur Verfügung, um die Paritätsmaschine an die Array-Organisation anzupassen. Die Maschine kann als einzelne Maschine konfiguriert werden, die an breiten Arrays mit einer Breite von bis zu 22+1 Laufwerken wirkt, oder als eine Gruppe von vier Maschinen, die mit Arrays mit Breiten von bis zu 4+1 Laufwerken arbeitet. Eine Zwischenbetriebsart sieht eine Gruppe von zwei Maschinen mit bis zu 10+1 Laufwerken vor.

Die Paritätsmaschine beinhaltet Exklusiv-ODER-Logik für RAID 5 und RAID 3 Paritätserzeugung/-prüfung sowie Bewegungs- und Nullprüfbetriebsarten.

Einer der wichtigsten Teile der Architektur ist der von ihr ermöglichte zeitunabhängige Betrieb. Es braucht auf keine Datenblöcke gleichzeitig für Paritätsberechnungen zugegriffen zu werden. Plattenvorgänge, die mehrere Plattenlaufwerke überspannen, können geplant und von jedem Gerät schnellstmöglich ausgeführt werden. Unverwandte Plattenvorgänge können auch dann fortgesetzt werden, wenn noch nicht alle Laufwerksvorgänge für eine einzelne Task abgeschlossen sind. Diese Unabhängigkeit verbessert die Leistung des langsamsten Teils des Systems, des Plattenlaufwerks. Sie vereinfacht auch die Software-Task des Verwaltens konkurrierender Hardware-Betriebsmittelanforderungen.

Die Erfindung ermöglicht zeitunabhängige Datenblockbewegungen, zeitunabhängige Paritätsberechnungen, in Warteschlangen befindliche RAID-Vorgänge (bis

zu 3 Tasks in einer Warteschlange), PCI-Übertragungsvorgänge bei voller Geschwindigkeit (4 Byte/Taktzyklus) sowie Paritätsberechnungen bei voller Geschwindigkeit (4 Byte/ Taktzyklus). Ferner ist die Architektur durch die Addition zusätzlicher RAID-Paritätsassist-Chips und zusätzlicher lokaler Speicherkapazität skalierbar.

So wird ersichtlich, dass die vorliegende Erfindung eine Plattenarray-Controller-Architektur mit unabhängiger Speicherstruktur zum Puffern von Daten sowie eine unabhängige Paritätsrechenmaschine bereitstellt, die einen Parallelbetrieb von Array-Tasks zulässt. Sie erlaubt Paritätsberechnungen und Speicherblockbewegungen, ohne dass die CPU oder ihr Speicher gestört wird, und befreit die CPU für die Verwaltung von Array-Tasksteuerung. Die offenbarte Array-Controller-Konfiguration erlaubt einen gleichzeitigen Ablauf von Datenblockbewegungen zwischen Speicher-E/A-Geräten und lokalem Speicher; Datenblockbewegungen zwischen Host-SCSI-Verbindungen und lokalem Speicher; Paritätsberechnungen; und normalen CPU-Speicherabrufen, in eine Warteschlange eingereihte Vorgänge für Blockbewegungen und in eine Warteschlange eingereihte Vorgänge für Paritätstasks.

Somit wird verständlich, dass die Erfindung ein Plattenarray-System bereitstellen kann, das Folgendes umfasst: einen ersten schnellen lokalen Bus; eine Schnittstellenschaltung, die den genannten ersten schnellen lokalen Bus mit dem genannten Host-Computersystem verbindet; eine Mehrzahl von Plattenlaufwerkselementen; wenigstens eine Schnittstellenschaltung, die die genannte Mehrzahl von Plattenlaufwerkselementen mit dem genannten ersten schnellen lokalen Bus verbindet; und einen Prozessor, der mit dem genannten ersten schnellen lokalen Bus verbunden ist, um den Betrieb der mit dem genannten ersten schnellen lokalen Bus verbundenen Komponenten zu steuern; mit den folgenden Verbesserungen, umfassend: eine Paritätsassist-Logikschaltung, die mit dem genannten ersten schnellen lokalen Bus verbunden ist; einen lokalen Speicher und einen zweiten schnellen lokalen Bus, der die genannte Paritätsassist-Logik mit dem genannten lokalen Speicher verbindet, wobei die genannte Paritätsassist-Logik Daten über den genannten zweiten schnellen lokalen Bus zu dem genannten lokalen Speicher bereitstellt, die für die Berechnung von Paritätsdaten bei Plattenarray-Schreibvorgängen benötigt werden, und die auf dem genannten lokalen Speicher gespeicherten Daten manipuliert, um Parität bei den genannten Array-Schreibvorgängen zu ermitteln.

Die Erfindung ist nicht auf die Einzelheiten der obigen Ausgestaltung begrenzt und es ist zu verstehen, dass verschiedene Änderungen innerhalb des Umfangs der beiliegenden Ansprüche möglich sind.

Ansprüche

EP 95 309 117.0

1. Plattenarray-Controller für den Betrieb mit einem Hostcomputer (12) und einer Mehrzahl von Plattenlaufwerken (114A-114E), umfassend einen ersten schnellen lokalen Bus (102), eine Schnittstelle (104), die den genannten ersten schnellen lokalen Bus (102) mit dem genannten Hostcomputer (12) verbindet, wenigstens eine Schnittstelle (112A-112E), die die genannte Mehrzahl von Plattenlaufwerken (114A-114E) mit dem genannten ersten schnellen lokalen Bus (102) verbindet, und einen Prozessor (106), der mit dem genannten ersten schnellen lokalen Bus (102) verbunden ist, um den Betrieb der mit dem genannten ersten schnellen lokalen Bus (102) verbundenen Komponenten zu steuern, gekennzeichnet durch Paritätsassist-Logik (132), die direkt mit dem genannten ersten schnellen lokalen Bus (102) verbunden ist, ein lokales Datenspeichermittel (136), das mit dem genannten ersten schnellen lokalen Bus (102) verbunden ist, und einen zweiten schnellen lokalen Bus (138), der die genannte Paritätsassist-Logik (132) mit dem genannten lokalen Datenspeichermittel (136) verbindet, wobei die genannte Paritätsassist-Logik (132) die Aufgabe hat, Daten über den genannten zweiten schnellen lokalen Bus (138) zu dem genannten lokalen Datenspeichermittel (136) zu übertragen, um Paritätsdaten während des Plattenarray-Schreibvorgangs zu bestimmen.

2. Vorrichtung nach Anspruch 1, bei der die genannte Paritätsassist-Logik (132) eine Paritätsassist-Maschine (132), eine Dualport-Speicherschnittstelle (134), die die genannte Paritätsassist-Maschine (132) mit dem genannten lokalen Datenspeichermittel (136) über den genannten zweiten schnellen lokalen Bus (138) verbindet, und eine Schnittstellenschaltung (130) zum Verbinden der genannten Dualport-Speicherschnittstelle (134) mit dem genannten ersten schnellen lokalen Bus (102) umfasst.

3. Vorrichtung nach Anspruch 1, bei der die genannte Paritätsassist-Logik (132) eine Exklusiv-ODER-Schaltung umfasst.

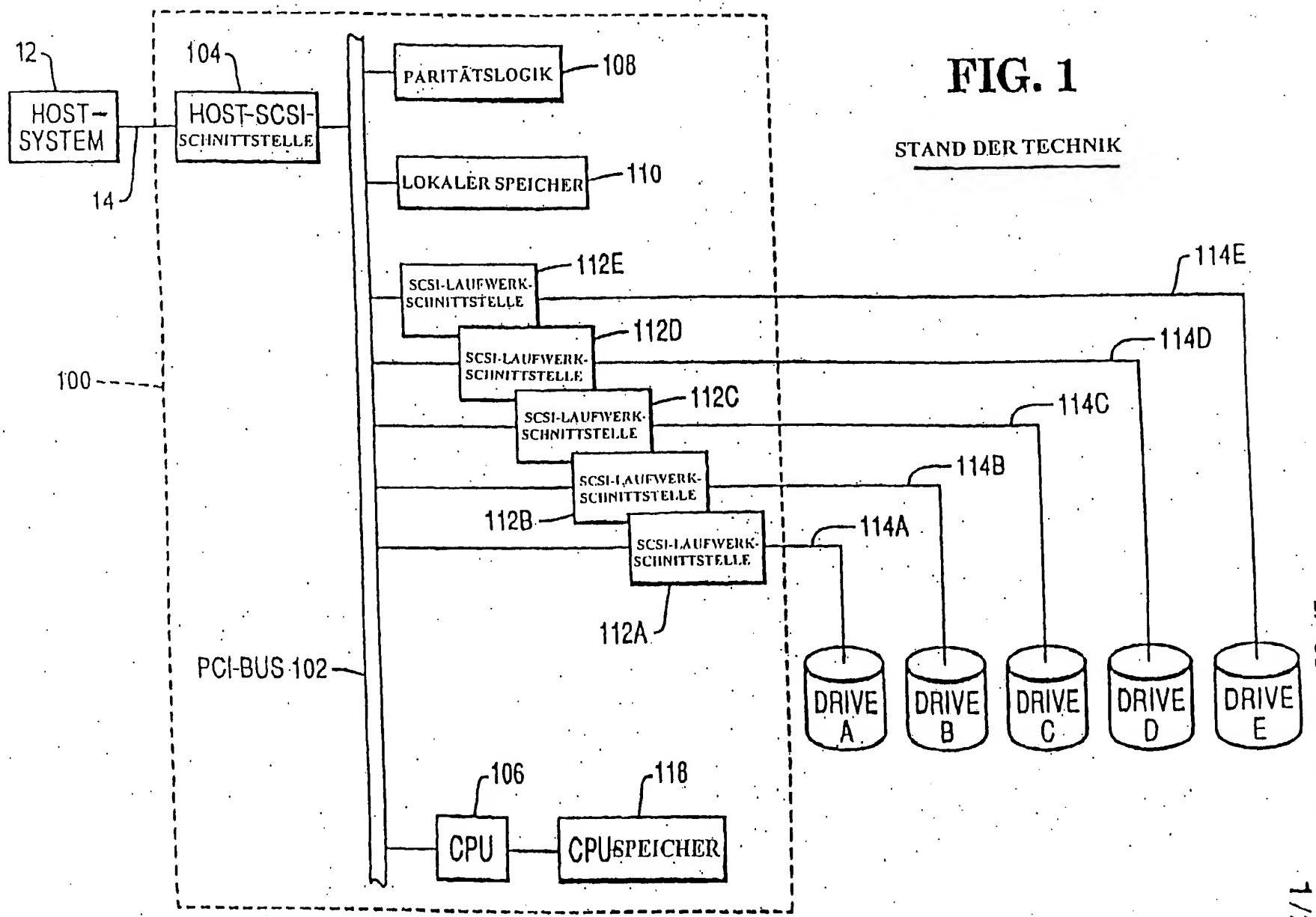

FIG. 1

STAND DER TECHNIK

EP 95 309 117.0

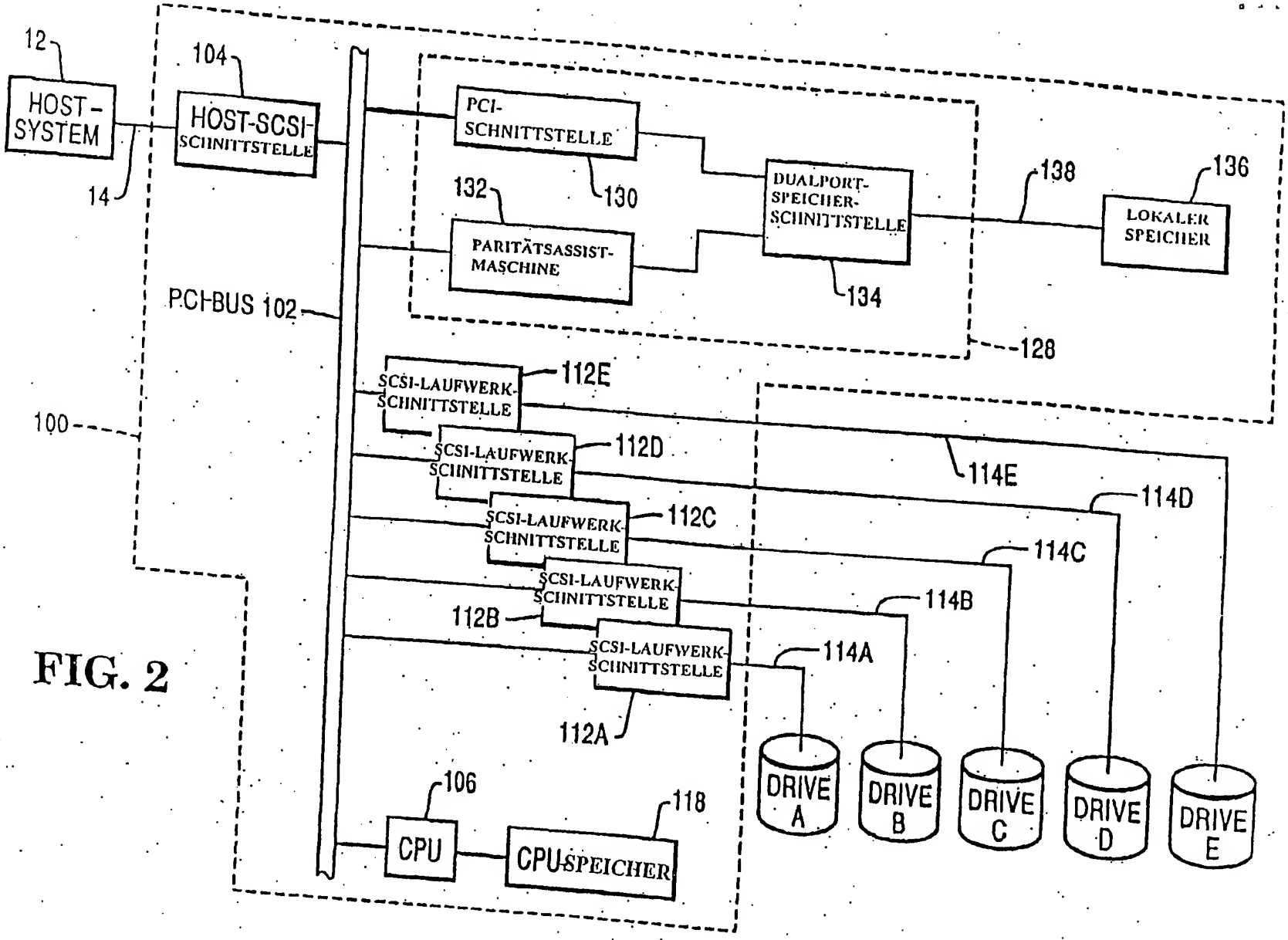
12
HOST-SYSTEM
14
104
HOST-SCSI-SCHNITTSTELLE
PCI-BUS 102
100
PCI-SCHNITTSTELLE
130
132
PARITÄTSASSIST-MASCHINE
DUALPORT-SPEICHER-SCHNITTSTELLE
134
138
136
LOKALER SPEICHER
128
SCSI-LAUFWERK-SCHNITTSTELLE
112E
112D
112C
112B
112A
114E
114D
114C
114B
114A
DRIVE A
DRIVE B
DRIVE C
DRIVE D
DRIVE E
106
CPU
118
CPU-SPEICHER

FIG. 2